

# 자연어 처리 기반 하이브리드 추천시스템

## Hybrid Recommendation System based on Natural Language Processing

이자영<sup>1</sup>, 임양미<sup>2\*</sup>

Ja-Young Lee<sup>1</sup>, Yang-Mi Lim<sup>2\*</sup>

### 요약

최근 온라인 유통업계에서는 판매자의 상품정보와 구매자의 리뷰 정보를 활용한 자연어처리 기반 상품 분석 및 추천 시스템이 활발하게 사용되고 있다. 본 연구에서는 특정 상품 양말의 데이터 특성과 구매자의 리뷰 데이터를 결합하여 활용한 자연어처리 기반의 추천 시스템을 개발하였다. 이 시스템은 양말에 대한 판매자의 상품 정보 중 성별, 길이, 패턴, 색상, 가격과 같은 요소를 분석하고, 이전 구매자의 리뷰 중 품질과 배송에 대한 평가를 기반으로 적합한 상품을 추천한다. 판매자의 상품 정보는 텍스트 데이터를 바탕으로, 리뷰 정보는 시퀀스 데이터를 바탕으로 학습되어 보다 정확한 추천을 제공한다. 본 연구에서는 키워드 간의 연관성을 분석하기 위해 TF-IDF 모델을 활용하고, 감정 분석을 위해 GRU 모델을 적용한 하이브리드 접근 방식을 제안한다. 양말 추천 시스템은 '도봉 양말협동조합'의 리뷰 데이터를 활용하여 지역 소상공인을 지원하기 위해 개발되었다. 이러한 시스템이 앞으로 도 양말과 같은 지역 상품 분야에서 더욱 널리 활용되어, 지역 경제 활성화에 기여할 수 있기를 기대한다.

핵심어 : 추천시스템, 감정분석, TF-IDF, GRU

### Abstract

In recent times, the online retail industry has seen active use of natural language processing (NLP)-based product analysis and recommendation systems that utilize seller-provided product information and customer review data. This study developed an NLP-based recommendation system by combining the data characteristics of a specific socks product, with customer review data. The system analyzes and categorizes seller-provided product information for socks (e.g., gender, length, pattern, color, price) and previous customer reviews (e.g., quality, delivery) to recommend the most suitable products. The seller's product information was processed through text-based learning, while the review data was processed using sequence-based learning, enhancing the accuracy of the recommendations. This study introduces a hybrid approach that employs the TF-IDF model for keyword association analysis and the GRU model for sentiment analysis. The sock recommendation system was developed using review data sourced from the

1 Department IT media Engineering, DukSung Women's University, Seoul, Korea [Undergraduate student]  
e-mail: wkud2019@duksung.ac.kr

2 Department IT media Engineering, DukSung Women's University, Seoul, Korea [Professor]  
e-mail: yosimi@duksung.ac.kr (Corresponding author)

\* 이 논문은 덕성여자대학교의 2023년도 교내연구비 지원을 받아 작성되었습니다.

Received(July 8, 2024), Review Result(1st: July 27, 2024), Accepted(August 9, 2024), Published(August 31, 2024)



© 2024 The Authors. Published by NCISS.  
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.  
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

'Dobong Sock Cooperative' to support local small businesses. It is anticipated that this system will be widely applied to regional products like socks in the future, thereby contributing to the revitalization of local economies.

Keyword : Recommendation system, Sentiment analysis, TF-IDF, GRU

## 1. 서론

최근 온라인 유통업계에서는 판매자의 상품정보와 구매자의 리뷰 정보를 이용하여 자연어 처리 기반 상품 분석 및 추천시스템을 도입하고 있다 [1]. 추천시스템은 상품 리뷰 데이터를 활용한 자연어 처리 기반 상품 분석, 감정 분석, 상품 평가를 통해 유사 상품 추천 서비스를 지원하고 있다 [2]. (주)네이버는 'AI 하이퍼클로바 (HyperCLOVA)'를 개발하였으며, 상품의 리뷰를 주요 키워드별로 평가하고 AI리뷰 서비스를 제공하고 있다. 최근 하이퍼클로바 X 모델로 확장된 이후에 상품 추천 뿐만 아니라 상품 리뷰 요약, 감정분석, 콘텐츠 분류, 문화적 뉘앙스까지 분석하여 다음 구매자에게 상품에 대한 요약된 정보도 제공한다 [3]. (주)크리마는 CREMA Review Solution으로 고객의 특성, 쇼핑 패턴 등을 분석하여 구매자의 맞춤형 고객 관리 시스템을 개발하여, 단순한 상품 추천이 아니라 상품의 사이즈 및 소비패턴에 맞는 상품까지 추천 정보로 제공하고 있다 [4]. (주)카카오페이지의 'AI 키토크' 시스템은 해당 작품의 맥락, 트렌드, 댓글, 다른 채널에서 사용자들의 반응과 이야기를 데이터를 학습하여 필터 형태로 검색 서비스를 제공한다 [5]. 이러한 검색 및 추천시스템은 구매자(사용자)가 다양한 상품을 지속적으로 구매하는데 도움을 준다.

본 연구에서는 특정 상품 양말을 지정하여 양말이 갖고 있는 특성을 고려한 추천시스템을 개발하였다. 양말은 옷의 하위 분야로 분류되어 양말만의 판매 촉진 플랫폼 개발이 이루어지지 않았었다. 특히 기존 의류와 다르게 천의 두께, 저렴한 가격, 특수 용도, 날씨, 계절, 미끄러움, 밴드 성능 등의 특징이 있으나 구체화된 정보가 반영되지 못하였다. 최근 추천서비스 기술을 활용하여 옷의 하위로 분류되었던 양말의 상품 정보를 양말의 사이즈, 감정 키워드 기반의 색상 추천이 가능하도록 개발하였다.

본 논문의 2장에서는 상품이 갖고 있는 데이터 특성을 반영하기 위한 추천시스템들을 설명하고, 3장에서는 양말 추천시스템 개발에 사용된 알고리즘들을 설명과 성능 테스트에 대한 설명을 한다. 4장은 개발된 양말 추천시스템 앱 구현 결과를 설명한다. 본 앱 개발은 지역사회 발전을 위해 소상공인 기업 지원 기술 제안으로 활용된 것으로 지역협업 연구 의의가 있다.

## 2. 자연어처리 기반 상품 추천 서비스

본 연구에서는 텍스트 기반과 감정분석을 이용한 추천시스템을 사용한다. 텍스트 기반 추천시스템은 콘텐츠기반 필터링, 협업필터링, 하이브리드 필터링으로 분류할 수 있다. 대표적으로 BoW

[6], TF-IDF [7], HF-IHU가 있다. 본 연구에 활용한 콘텐츠 기반 필터링인 TF-IDF(Term Frequency-Inverse Document Frequency)는 특정 문서 내에서 키워드 생성을 위해 가중치를 계산하는 방법이다. 문서의 키워드를 추출하고 문서들 사이의 유사도를 계산, 검색 결과의 중요도를 정하고 추천하는 단계를 거친다 [8][9]. 리뷰데이터에 대한 감정분석에는 RNN(Recurrent Neural Networks)기반의 딥러닝이 많이 사용되어 왔다. 하지만, RNN은 기울기 소실 문제가 있어, LSTM(long short term memory networks)으로 문제를 완화하였다. LSTM 또한 학습할 매개변수가 많은 단점을 보완하기 위해 매개변수를 적게하고 연산을 빠르게 하기 위한 GRU가 사용되고 있다 [10][11].

[표 1]은 자연어처리 기반의 추천시스템에서 사용된 알고리즘을 정리한 것이다. 첫 번째 김민우 연구는 다양한 장르의 책 추천을 위해 책의 소개 글에 나오는 단어 빈도를 TF-IDF 행렬을 사용하여, 사용자가 입력한 단어들의 유사도를 Word2vec을 이용하여 구하고 텍스트로 변환한 후, 코사인 유사도로 구매자의 취향에 맞는 책(상품)을 추천하였다 [12]. 두 번째는 텍스트 마이닝의 감성 분석 기법을 통해 감성 수치를 산출하였다. 영화 리뷰를 대상으로 하여 라이노(RHINO) 라이브러리를 이용한 전처리와 감성 사전을 구축하고 이를 기반으로 리뷰의 감성 점수를 산출하였다 [13]. 세 번째는 기계학습 기반의 포지션 임베딩을 이용하여 문장에서 한 쌍의 단어 관계를 정의하고 LSTM을 훈련해 BERT 기반 감성분석을 활용한 추천시스템을 개발하였다.

제안한 양말 추천시스템은 구매자의 리뷰 정보인 시퀀스 데이터와 판매자의 제품 정보인 메타 데이터를 고려하여, LSTM과 GRU 모델로 학습하는 하이브리드 방식을 제안한다.

[표 1] TF-IDF, RHINO, LSTM 비교

[Table 1] Comparison of TF-IDF, RHINO, LSTM

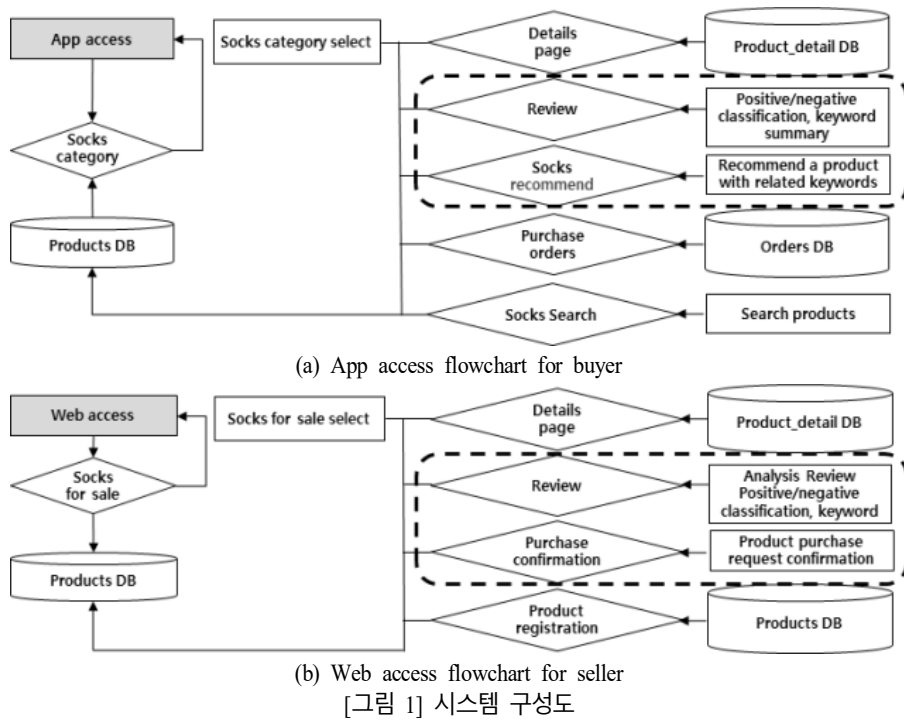
No	Title	NLP	Research purpose
1	A Study on the Book Recommendation System Using Natural Language Processing	TF-IDF, Word2vec	Compare word similarity
2	How to improve the accuracy of recommendation systems : Combining ratings and review texts sentiment scores	RHINO, A Logistic Regression	Recommendation using emotional analysis
3	Recommender system using BERT sentiment analysis	Bert (LSTM)	Recommendation using emotional analysis by BERT
4	Proposed research	TF-IDF, Word2vec, LSTM, GRU	Recommendation using word similarity, Review emotional analysis

### 3. 양말 추천서비스 시스템 개발

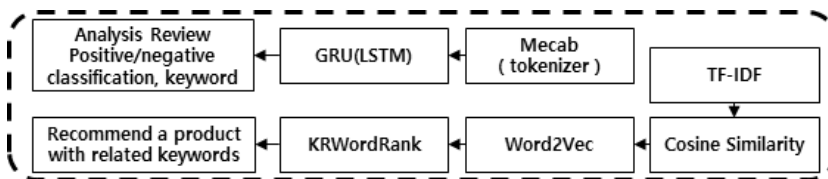
#### 3.1 시스템 설계

제안한 추천시스템은 양말에 대한 판매자의 상품 정보(성별, 길이, 패턴, 색상, 가격)와 이전 구

매자의 리뷰 정보(품질, 배송) 데이터 중심으로 분석, 분류, 추천하고 있다. 감정분석을 위해서는 상품 리뷰 만족도를 표로 시각화하여 긍정/부정 리뷰를 확인할 수 있도록 하였고 학습 모델로 GRU 모델을 사용하였다. GRU 모델은 RNN의 장기 의존성문제(long-term dependencies)와 LSTM의 복잡한 수식 해결을 위해 개발된 딥러닝 모델 중의 하나이다. [그림 1]은 시스템 구성도이다. 앱은 주로 구매자 중심으로 데이터의 흐름을 설명하고 있고, Web는 판매자가 올리는 데이터와 상품리뷰 데이터의 흐름을 설명한다. 점선 박스는 양말 추천을 위해 사용된 단어 유사도 검출과 구매자의 리뷰를 기반으로 감정 분석을 위해 사용된 GRU 모델 사용에 대한 것으로 [그림 2]에서 상세 흐름을 설명한다.



[Fig. 1] System work flowchart



[Fig. 2] Proposed GRU and TF-IDF hybrid natural language processing model

[그림 2]는 상품 추천을 위해 판매자가 제공한 양말 정보, 구매자가 제공한 리뷰를 통해 양말 상품을 키워드 한 정보, 리뷰를 감성 분석에 따라 요약한 양말 상품 정보를 세분화하여 Text encoder 모델로 사용한다. 판매자가 제공한 양말 정보는 사용자가 양말의 용도인 성별, 길이, 패턴, 색상들의 메타데이터로 입력 받은 것을 초기 협업 필터링에 구성된다. 각 양말 상품의 리뷰와 유사도가 높은 키워드 3가지 제시하며 한 키워드를 선택할 때 키워드를 반영한 협업 필터링이 구성된다.

### 3.2 제안한 GRU와 TF-IDF 기반의 추천시스템

양말 추천을 위한 데이터 수집은 ‘도봉 양말협동조합’의 양말 리뷰 데이터를 활용하였다. 상품 추천을 위한 유사도 값을 구하는데 있어서 데이터 셋 구축, 데이터 전처리, 코사인 유사도 방식을 활용하였다. 첫 번째 과정인 데이터 셋 구축과 추천 키워드가 생성은 26,042개의 리뷰 텍스트 데이터 셋의 중복 리뷰 제거와 상품과 연관성이 낮고 양말의 특징을 보유하지 않는 ‘너무’, ‘자주’와 같은 의미 없는 파생어 단어를 제거하는 전처리 과정을 거친 후 데이터 셋으로 구성하였다. 제거된 데이터 셋을 KRWordRank으로 키워드 추출하였고, 상품 추천키워드로서 적용할 수 없거나 실용이 없는 단어들을 불용어 처리하였다. KRWordRank는 텍스트의 tokenizer를 이용하지 않으면서도 한국어 특징을 반영하고 비지도 학습 기반으로 키워드 추출하는 라이브러리이다 [14]. 최종적으로 [그림 2]와 같이 KRWordRank를 통해 가장 양말과 연관성 있는 특징 단어 30개를 형용사로 말뭉치화한 후 양말 추천키워드로 표현하였다.

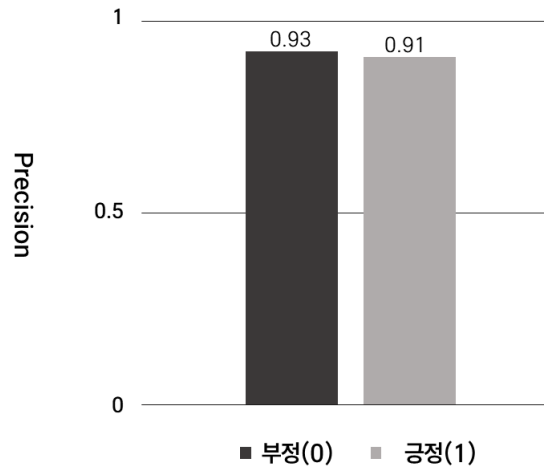
두 번째 과정은 자연어 처리의 전처리로 상품 리뷰 내에서 출현한 단어 빈도를 구하여 TF-IDF 행렬을 수행하였다. TF-IDF은 특정 단어가 문서 내에 얼마나 자주 등장하는지 나타내는 TF(term frequency)값과 다른 문서에서 특정 단어가 얼마나 자주 등장하는지 나타내는 DF(document fequency)의 역수 값인 IDF로 행렬을 나타낸다 [15]. cold-start 문제를 해결과 유사도를 계산을 위한 워드 임베딩 행렬로는 Word2Vec과 양말 추천 키워드와 유사한 단어 6개를 추출하여 텍스트 변환 과정을 실행하였다. Word2Vec은 추론 기반 기법으로 문맥적 의미를 보존하여 단어를 벡터로 표현한다. 또한, 비슷한 단어와 단어 벡터 사이의 거리를 통해 유사한 단어를 제공 역할을 한다 [16].

마지막으로 최적의 유사도 값을 구하기 위해 한 키워드의 텍스트를 TF-IDF 행렬과 코사인 유사도 값을 구하였다. 식 (1)의 코사인 유사도는 문서와 문서 사이의 유사도를 구하여 두 문서가 얼마나 유사한지를 나타낸다. 전체 30개의 키워드의 TF-IDF 행렬과 유사도 계산 후 유사도가 높은 값인 1과 가까운 키워드 3개를 적용하여 상품을 대표하는 양말 추천키워드로 적용하였다.

$$\text{cosine similarity} = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (1)$$

### 4. 구현 결과

[그림 3]는 감정 모델의 학습 정확도를 나타낸다. 감정분석을 위한 긍정/부정 분류 모델의 정확도 향상을 위해 LSTM보다 속도가 빠른 GRU 모델로 학습하였으며, 부정의 경우 정확도는 0.93, 긍정은 0.91로 측정되었다.



[그림 3] 감정분석 모델 정확도

[Fig. 3] Emotion Analysis Model Accuracy

[표 2]는 상품과 연관된 30개의 키워드 생성 결과를 보여준다. 상품리뷰 키워드 과정에서 불용어 리스트를 추가하여, 26,042개의 리뷰 텍스트 데이터셋 키워드화하여 상품과 가장 연관성 있으며 특징을 가진 단어 30개의 키워드를 추출하였다. 추출한 단어 30개를 상품과 연관하여 형용사로 말뭉치화한 후 상품 추천 키워드로 적용하였다.

[표 2] 키워드 생성 단어 리스트

[Table 2] Keyword creation word list

index	keyword	frequency	index	keyword	frequency
0	양말이두툽한	158	15	대량구매하는	177
1	양말이저렴한	169	16	가족에게좋은	89
2	교복에어울리는	147	17	재구매예정인	199
3	더운날좋은	144	18	운동에좋은	57
4	추운날에좋은	135	19	화면과똑같은	89
5	봄에어울리는	121	20	양말이깔끔한	197
6	여름에어울리는	82	21	두께가얇은	84

7	가을에어울리는	41	22	늘어나지않는	176
8	미끄러움이 적은	110	23	구멍이안나는	153
9	촉감이부드러운	133	24	재질이좋은	216
10	밴드가쫄쫄한	148	25	착용감이좋은	210
11	서비스가좋은	176	26	땀흡수가좋은	56
12	샤은품이푸짐한	200	27	두께가적당한	74
13	양말이편한	189	28	두께가두툼한	86
14	아이들에좋은	87	29	활용하기좋은	164

[표 3]은 상품 추천 키워드는 한국어로 사전 학습된 Word2Vec을 이용하여 각각의 상품 추천 키워드와 유사한 단어 6개를 찾아서 이를 하나의 텍스트들로 만들었다. 한 키워드의 텍스트를 TF-IDF 행렬들로 만들어 한 상품의 리뷰 TF-IDF 행렬과 코사인 유사도 계산을 한다. 전체 30개의 키워드의 TF-IDF 행렬과 유사도 계산 후 유사도 1과 가까운 키워드 3개를, 상품을 대표하는 키워드로 적용한 결과를 보여준다.

[표 3] 키워드와 상품 유사도

[Table 3] Comparison of keyword similarity with product

index	keyword	similarity
0	대량구매하는	0.616111249
1	가족에게좋은	0.56794672
2	재구매예정인	0.363401316
3	서비스가좋은	0.356810212
4	샤은품이푸짐한	0.244803934
5	양말이편한	0.244287693
6	활용하기좋은	0.239405444
7	미끄러움이 적은	0.237166955
8	촉감이부드러운	0.234643889
9	밴드가쫄쫄한	0.22805799

[그림 4]는 학습된 감정분석 모델을 적용하여 테스트한 결과이다. 리뷰 내용 중 ‘재구매각’ ‘좋아요’가 들어간 리뷰들에서 정확하게 긍정 리뷰로 예측한 것을 알 수 있다.

[그림 5]는 제안한 양말 추천서비스 앱을 보여준다. (a)는 성별, 길이(양말사이즈), 무늬, 용도에 맞게 추천된 리스트 화면이다. (b)는 구매자가 선호한 상품 선택 화면이다. (c)는 구매자가 결정한 상품을 결제하는 장바구니 화면이다.

평점	리뷰	리뷰분석
0	5	한달사용기가성비짱입니다재구매각 98.45% 확률로 긍정 리뷰입니다.
1	5	좋아요에민보스 깔칠중3 아들이 괜찮다고하면 괜찮은거예요 94.46% 확률로 긍정 리뷰입니다.
2	5	디자인도 예쁘고 저렴한데 재질도 튼튼해요대만족입니다게다가 첫구매라 그런지 서비스 양... 99.18% 확률로 긍정 리뷰입니다.
3	5	랜덤으로 했더니 색깔 중복없이 주셨고 20개이상사니 서비스도 주셨네요 양말 질이 넘... 93.27% 확률로 긍정 리뷰입니다.
4	5	소소하게 귀여워요 잘샀어요 91.64% 확률로 긍정 리뷰입니다.
5	5	중학생 아이가 신을 양말 두가지 버전으로 구입했어요 이제 한동안 양말 걱정 없겠네요 98.56% 확률로 긍정 리뷰입니다.
6	5	제품 깔끔하고 맘에 들어요 92.67% 확률로 긍정 리뷰입니다.
7	5	재구매택배비닐이 뜯어 있어서 포장 잘해주세요 72.19% 확률로 부정 리뷰입니다.
8	5	양말도 이쁘고 사은품으로 주신 양말도 좋아요 97.89% 확률로 긍정 리뷰입니다.
9	5	저렴히 잘샀것 같아요ㅋㅋ잘 신을게용ㅋㅋ 98.84% 확률로 긍정 리뷰입니다.

[그림 4] 모델 테스트

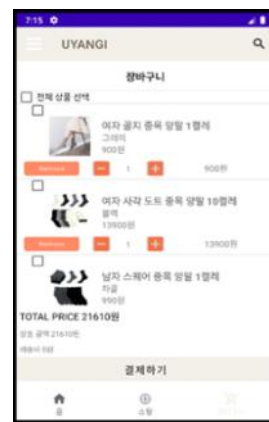
[Fig. 4] Model Test



(a) 상품 추천 리스트  
(a) Product recommendation list



(b) 구매자 선호 상품 선택  
(b) Select buyer's preferred product



(c) 구매 결정 장바구니  
(c) purchase decision shopping cart

[그림 5] 추천상품 구매 과정

[Fig. 5] Recommended product purchase process

## 5. 결론

본 논문은 판매자가 입력한 상품 정보와 구매자가 입력한 상품 리뷰 데이터를 기반으로 하여 양말 상품을 추천하는 시스템을 개발하였다. 판매자가 입력한 상품정보인 텍스트 기반 데이터와 구매자가 입력한 리뷰데이터인 시퀀스 데이터에 대한 효율적 처리를 위해 LSTM과 GRU 모델로 학습하는 하이브리드 방식을 제안하였다. 리뷰데이터를 활용한 감정분석은 온라인 쇼핑몰에서 활발하게 활용되고 있는 마케팅 기법 중의 하나이다. 여러 제품이 모두 담겨 있는 쇼핑몰에서 리뷰

데이터는 상품 카테고리 별로 다른 특징을 갖고 있다. 따라서 특정한 상품에 대한 추천을 정확하게 정보를 받기에는 문제점이 많다. 본 연구는 제품에 대한 정보를 판매자가 등록한 상품정보를 텍스트 정보로 학습하고, 리뷰데이터를 입력한 모델과 하이브리드 방식으로 학습하여 감정분석 모델을 제안하였다. 이러한 방식은 지역 소상공인이 개발한 다양한 상품을 구매자(사용자)가 상품 정보를 추천 받는 데 있어 효율적으로 적용될 것이라고 예상된다.

## References

- [1] M. Y. Kim, et al 395 authors, "HyperCLOVA X Technical Report", published Cornell University, April 2024, doi: 10.48550/arXiv.2404.01954.
- [2] S. J. Kim, Y. S. Kim, "Food Recommendation System Using BigData Based on Scoring Taste Adjectives Sentiment Dictionary", *Journal of Next-generation Convergence Information Services Technology*, vol. 5, no. 1, January 2016, pp. 9-16, doi: 10.34163/jkits.2019.14.4.002.
- [3] CLOVA X, "Naver Generative AI", clova-x.naver.com, <https://clova-x.naver.com/welcome>, (accessed May. 01, 2024).
- [4] CREMA, "Start CREMA Widget", dev.cre.ma, <https://dev.cre.ma/crema-widgets/getting-started>, (accessed May. 01, 2024)
- [5] Keytalk AI, "Transform Your Business with Semantic AI Prompts", keytalkai.com, <https://www.keytalkai.com>, (accessed May. 01, 2024)
- [6] H. Yin, W. Wang, L. Chen, X. Du, Q. V. H. Nguyen, Z. Huang, "Mobi-Sage-Rs: A Sparse Additive Generative Model-Based Mobile Application Recommender System", *Knowledge-Based Systems*, vol. 157, October 2018, pp. 68-80 doi: 10.1016/j.knsys.2018.05.028.
- [7] E. Zangerle, W. Gassler, G. Specht, "On the Impact of Text Similarity Functions on Hashtag Recommendations in Microblogging Environments", *Social network analysis and mining*, vol. 3, no. 4, March 2013, pp. 889-898, doi: 10.1007/s13278-013-0108-x.
- [8] H. Y. Park, K. J. Kim, "Recommender system using BERT sentiment analysis", *Journal of Intelligence and Information Systems*, vol. 27, no. 2, June 2021, pp. 1-15, doi: 10.13088/jiis.2021.27.2.001.
- [9] A. Da'u, N. Salim, "Sentiment-aware deep recommender system with neural attention networks", *IEEE Access*, vol. 7, March 2019, pp. 45472-45484, doi: 10.1109/ACCESS.2019.2907729.
- [10] H. H. Hwang, K. C. Lee, J. Y. Yu, Y. H. Lee, "Multimodal Sentiment Analysis Using Review Data and Product Information", *The Journal of Society for e-Business Studies*, vol. 27, no. 1, February 2022, pp. 15-28. doi : 10.7838/jsebs.2022.27.1.015.
- [11] D. Y. Lee, Y. H. Yoo, H. S. Lim, "Bi-directional LSTM-CNN-CRF for Korean Named Entity Recognition System with Feature Augmentation", *Journal of the Korea Convergence Society*, vol. 8, no. 12, December 2017, pp. 55-62, doi: 10.15207/JKCS.2017.8.12.055.
- [12] M. W. Kim, S. H. Kim, D. Y. Park. "A Study on the Book Recommendation System Using Natural Language Processing", *Proceedings of Symposium of the Korean Institute of Communications and*

- Information Sciences, 03-05 Winter 2021, Yongpyong Korea, pp. 869-870.
- [13] J. Y. Hyun, S. Y. Ryu, S. Y. Lee, "How to improve the accuracy of recommendation systems : Combining ratings and review texts sentiment scores", *Journal of Intelligence and Information Systems*, vol. 25, no. 1, March 2019, pp. 219-239, doi: 10.13088/jiis.2019.25.1.219.
- [14] H. J. Kim, S. Z. Cho, P. S. Kang, "KR-WordRank : An Unsupervised Korean Word Extraction Method Based on WordRank", *Journal of the Korean Institute of Industrial Engineers*, vol. 40, no. 1, February 2014, pp. 18-33, doi: 10.7232/JKIIIE.2014.40.1.018.
- [15] L. P. Jing, H. K. Huang, H. -B. Shi, "Improved feature selection approach TFIDF in text mining", *Proceedings. International Conference on Machine Learning and Cybernetics*, vol. 2, 04-05 November 2002, Beijing, China, pp. 944-946. doi: 10.1109/ICMLC.2002.1174522
- [16] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space", arXiv:1301.3781v3, *Computation and Language*, published Cornell University, January 2013, doi: 10.48550/arXiv.1301.3781.